Systems of Linear Equations

0.1 Definitions

Recall that if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times p}$, then the **augmented matrix** $[A \mid B] \in \mathbb{R}^{m\times n+p}$ is the matrix $[A\ B]$, that is the matrix whose first n columns are the columns of A, and whose last p columns are the columns of B. Typically we consider $B = \ \in \mathbb{R}^{m\times 1} \simeq \mathbb{R}^m$, a column vector.

We also recall that a matrix $A \in \mathbb{R}^{m\times n}$ is said to be in **reduced row echelon form** if, counting from the topmost row to the bottom-most,

1. any row containing a nonzero entry precedes any row in which all the entries are zero (if any)
2. the first nonzero entry in each row is the only nonzero entry in its column
3. the first nonzero entry in each row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row

Example 0.1 *The following matrices are* not *in reduced echelon form because they all fail some part of 3 (the first one also fails 2):*

$$\begin{pmatrix}1&1&0\\0&1&0\\1&0&1\end{pmatrix}\qquad\begin{pmatrix}0&1&0&2\\1&0&0&1\\0&0&1&1\end{pmatrix}\qquad\begin{pmatrix}2&0&0\\0&1&0\end{pmatrix}$$

A matrix that is in reduced row echelon form is:

$$\begin{pmatrix}1&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}$$

■

A **system of** m **linear equations in** n **unknowns** is a set of m equations, numbered from 1 to m going down, each in n variables x_i which are multiplied by **coefficients** $a_{ij} \in F$, whose sum equals some $b_j \in \mathbb{R}$:

$$(S)\qquad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

If we condense this to matrix notation by writing $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{b} = (b_1, \ldots, b_m)$ and $A \in \mathbb{R}^{m\times n}$, the **coefficient matrix** of the system, the matrix whose elements are the coefficients a_{ij} of the variables in (S), then we can write (S) as

$$(S)\qquad A\mathbf{x} = \mathbf{b}$$

noting, of course, that $\mathbf{b}$ and $\mathbf{x}$ are to be treated as column vectors here by associating $\mathbb{R}^n$ with $\mathbb{R}^{n\times 1}$. If $\mathbf{b} = \mathbf{0}$ the system (S) is said to be **homogeneous**, while if $\mathbf{b} \neq \mathbf{0}$ it is said to be **nonhomogeneous**. Every nonhomogeneous system $A\mathbf{x} = \mathbf{b}$ has an **associated** or **corresponding homogeneous system** $A\mathbf{x} = \mathbf{0}$. Furthermore, each system $A\mathbf{x} = \mathbf{b}$, homogeneous or not, has an **associated** or **corresponding augmented matrix** is the $[A \mid b] \in \mathbb{R}^{m\times n+1}$.

A **solution** to a system of linear equations $A\mathbf{x} = \mathbf{b}$ is an n-tuple $\mathbf{s} = (s_1, \ldots, s_n) \in \mathbb{R}^n$ satisfying $A\mathbf{s} = \mathbf{b}$. The **solution set** of $A\mathbf{x} = \mathbf{b}$ is denoted here by K. A system is either **consistent**, by which we mean $K \neq \varnothing$, or **inconsistent**, by which we mean $K = \varnothing$. Two systems of linear equations are called **equivalent** if they have the same solution set. For example the systems $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{c}$, where $[B \mid \mathbf{c}] = \operatorname{rref}([A \mid \mathbf{b}])$ are equivalent (we prove this below).

0.2 Preliminaries

Remark 0.2 *Note that we here use a different (and more standard) definition of* **rank of a matrix***, namely we define* $\operatorname{rank} A$ *to be the dimension of the image space of* A*,* $\operatorname{rank} A := \dim(\operatorname{im} A)$*. We will see below that this definition is equivalent to the one in Bretscher's* Linear Algebra With Applications *(namely, the number of leading* 1s *in* $\operatorname{rref}(A)$*).* ■

Theorem 0.3 *If* $A \in \mathbb{R}^{m\times n}$*,* $P \in \mathbb{R}^{m\times m}$ *and* $Q \in \mathbb{R}^{n\times n}$*, with* P *and* Q *invertible, then*

(1) $\operatorname{rank}(AQ) = \operatorname{rank}(A)$

(2) $\operatorname{rank}(PA) = \operatorname{rank}(A)$

(3) $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$

Proof: (1) If Q is invertible then the associated linear map T_Q is invertible, and so bijective, so that $\operatorname{im} T_Q = T_Q(\mathbb{R}^n) = \mathbb{R}^n$. Consequently

$$\operatorname{im}(T_{AQ}) = \operatorname{im}(T_A \circ T_Q) = T_A\big(\operatorname{im}(T_Q)\big) = T_A(\mathbb{R}^n) = \operatorname{im}(T_A)$$

so that

$$\operatorname{rank}(AQ) = \dim\big(\operatorname{im}(T_{AQ})\big) = \dim\big(\operatorname{im}(T_A)\big) = \operatorname{rank}(A)$$

(2) Again, since T_P is invertible, and hence bijective, because P is, we must have

$$\dim\big(\operatorname{im}(T_P \circ T_A))\big) = \dim(\operatorname{im}(T_A))$$

Thus,

$$\begin{aligned}\operatorname{rank}(AQ) = \dim\big(\operatorname{im}(T_{AQ})\big) = \dim\big(\operatorname{im}(T_P \circ T_A)\big)& \\ = \dim\big(\operatorname{im}(T_P \circ T_A)\big) = \dim(\operatorname{im}(T_A)) = \operatorname{rank}(A)&\end{aligned}$$

(3) This is just a combination of (1) and (2): $\operatorname{rank}(PAQ) = \operatorname{rank}(AQ) = \operatorname{rank}(A)$. ■

Corollary 0.4 *Elementary row and column operations on a matrix are rank-preserving.*

Proof: If B is obtained from A by an elementary row operation, there exists an elementary matrix E such that $B = EA$. Since elementary matrices are invertible, the previous theorem implies $\operatorname{rank}(B) = \operatorname{rank}(EA) = \operatorname{rank}(A)$. A similar argument applies to column operations. ■

Theorem 0.5 *A linear transformation* $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ *is injective iff* $\ker(T) = \{0\}$*.*

Proof: If T is injective and $\mathbf{x} \in \ker(T)$, then $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$, so that $\mathbf{x} = \mathbf{0}$, whence $\ker(T) = \{\mathbf{0}\}$. Conversely, if $\ker(T) = \{\mathbf{0}\}$ and $T(\mathbf{x}) = T(\mathbf{y})$, then,

$$\mathbf{0} = T(\mathbf{x}) - T(\mathbf{y}) = T(\mathbf{x} - \mathbf{y}) \quad\Longrightarrow\quad \mathbf{x} - \mathbf{y} = \mathbf{0}$$

or $\mathbf{x} = \mathbf{y}$, and so T is injective. ■

Theorem 0.6 *A linear transformation $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is injective iff it carries linearly independent sets into linearly independent sets.*

Proof: If T is injective, then $\ker T = \{\mathbf{0}\}$, and if $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent, then for all $a_1, \ldots, a_k \in \mathbb{R}$ we have $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0} \implies a_1 = \cdots = a_k = 0$. Consequently, if

$$a_1T(\mathbf{v}_1) + \cdots + a_kT(\mathbf{v}_k) = \mathbf{0}$$

then, since $a_1T(\mathbf{v}_1)+\cdots+a_kT(\mathbf{v}_k) = T(a_1\mathbf{v}_1+\cdots+a_k\mathbf{v}_k)$, we must have $a_1\mathbf{v}_1+\cdots+a_k\mathbf{v}_k \in \ker T$, or $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$, and so

$$a_1 = \cdots = a_k = 0$$

whence $T(v_1), \ldots, T(v_n) \in \mathbb{R}^m$ are linearly independent. Conversely, if T carries linearly independent sets into linearly independent sets, let $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and suppose $T(\mathbf{u}) = T(\mathbf{v})$ for some $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Since $\mathbf{u} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ and $\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$ for unique $a_i, b_i \in \mathbb{R}$, we have

$$\begin{aligned} \mathbf{0} = T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v}) &= T\big((a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n\big) \\ &= (a_1 - b_1)T(\mathbf{v}_1) + \cdots + (a_n - b_n)T(\mathbf{v}_n) \end{aligned}$$

so that, by the linear independece of $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$, we have $a_i - b_i = 0$ for all i, and so $a_i = b_i$ for all i, and so $\mathbf{u} = \mathbf{v}$ by the uniqueness of expressions of vectors as linear combinations of basis vectors. Thus, $T(\mathbf{u}) = T(\mathbf{v}) \implies \mathbf{u} = \mathbf{v}$, which shows that T is injective. ■

0.3 Important Results

Theorem 0.7 *The solution set K of any system $A\mathbf{x} = \mathbf{b}$ of m linear equations in n unknowns is an affine space, namely a coset of $\ker(T_A)$ represented by a particular solution $\mathbf{s} \in \mathbb{R}^n$:*

$$K = \mathbf{s} + \ker(T_A) \tag{0.1}$$

Proof: If $\mathbf{s}, \mathbf{w} \in K$, then

$$A(\mathbf{s} - \mathbf{w}) = A\mathbf{s} - A\mathbf{w} = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

so that $\mathbf{s} - \mathbf{w} \in \ker(T_A)$. Now, let $\mathbf{k} = \mathbf{s} - \mathbf{w} \in \ker(T_A)$. Then,

$$\mathbf{w} = \mathbf{s} + \mathbf{k} \in \mathbf{s} + \ker(T_A)$$

Hence $K \subseteq \mathbf{s} + \ker(T_A)$. To show the converse inclusion, suppose $\mathbf{w} \in \mathbf{s} + \ker(T_A)$. Then $\mathbf{w} = \mathbf{s} + \mathbf{k}$ for some $\mathbf{k} \in \ker(T_A)$. But then

$$A\mathbf{w} = A(\mathbf{s} + \mathbf{k}) = A\mathbf{s} + A\mathbf{k} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

so $\mathbf{w} \in K$, and $\mathbf{s} + \ker(T_A) \subseteq K$. Thus, $K = \mathbf{s} + \ker(T_A)$. ■

Theorem 0.8 *Let $A\mathbf{x} = \mathbf{b}$ be a system of n linear equations in n unknowns. The system has exactly one solution, $A^{-1}\mathbf{b}$, iff A is invertible.*

Proof: If A is invertible, substituting $A^{-1}\mathbf{b}$ into the equation gives

$$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$$

so it is a solution. If $\mathbf{s}$ is any other solution, then $A\mathbf{s} = \mathbf{b}$, and consequently $\mathbf{s} = A^{-1}\mathbf{b}$, so the solution is unique. Conversely, if the system has exactly one solution $\mathbf{s}$, then by the previous theorem $K = \mathbf{s} + \ker(T_A) = \{\mathbf{s}\}$, so $\ker(T_A) = \{\mathbf{0}\}$, and T_A is injective. But it is also onto, because $T_A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ takes linearly independent sets into linearly independent sets: explicitly, it takes a basis $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ to a basis $T_A(\beta) = \{T_A(\mathbf{v}_1), \ldots, T_A(\mathbf{v}_n)\}$ (because if $T(\beta)$ is linearly independent, it is a basis by virtue of having n elements). Because it is a basis, $T_A(\beta)$ spans $\mathbb{R}^n$, so that if $\mathbf{v} \in \mathbb{R}^n$, there are $a_1, \ldots, a_n \in \mathbb{R}$ such that

$$\mathbf{v} = a_1 T_A(\mathbf{v}_1) + \cdots + a_n T_A(\mathbf{v}_n) = T_A(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n)$$

Letting $\mathbf{u} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n\mathbb{R}^n$ shows that $T_A(\mathbf{u}) = \mathbf{v}$, so T_A, and therefore A, is surjective, and consequently invertible. ■

Theorem 0.9 *A system of linear equations $A\mathbf{x} = \mathbf{b}$ is consistent iff* $\operatorname{rank} A = \operatorname{rank}[A|\mathbf{b}]$.

Proof: Obviously $A\mathbf{x} = \mathbf{b}$ is consistent iff $\mathbf{b} \in \operatorname{im} T_A$. But in this case

$$\operatorname{im} T_A = \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \operatorname{span}(\mathbf{a}_1, \ldots, \mathbf{a}_n, \mathbf{b}) = \operatorname{im} T_{[A|\mathbf{b}]}$$

where $\mathbf{a}_i$ are the columns of A. Therefore, $A\mathbf{x} = \mathbf{b}$ is consistent iff

$$\operatorname{rank} A = \dim\big(\operatorname{im} T_A\big) = \dim\big(\operatorname{im} T_{(A|\mathbf{b})}\big) = \operatorname{rank}\big([A|\mathbf{b}]\big)$$ ■

Corollary 0.10 *If $A\mathbf{x} = \mathbf{b}$ is a system of m linear equations in n unknowns and it's augmented matrix $[A|\mathbf{b}]$ is transformed into a reduced row echelon matrix $[A'|\mathbf{b}']$ by a finite sequence of elementary row operations, then*

(1) $A\mathbf{x} = \mathbf{b}$ *is inconsistent iff* $\operatorname{rank}(A') \neq \operatorname{rank}[A'|\mathbf{b}']$ *iff* $[A'|\mathbf{b}']$ *contains a row in which the only nonzero entry lies in the last column, the* $\mathbf{b}'$ *column.*

(2) $A\mathbf{x} = \mathbf{b}$ *is consistent iff* $[A'|\mathbf{b}']$ *contains no row in which the only nonzero entry lies in the last column.*

Proof: If $\operatorname{rank} A' \neq \operatorname{rank}[A'|\mathbf{b}']$, then $\operatorname{rank}(A') < \operatorname{rank}[A'|\mathbf{b}']$, since we could consider A' as equal to $[A'|\mathbf{0}]$, and if this matrix has r linearly independent rows, or rank r, so does A'. Whence if $\operatorname{rank}[A'|\mathbf{b}'] \neq \operatorname{rank}[A'|\mathbf{0}] = \operatorname{rank} A'$, it is because $\mathbf{b}'$ contains some nonzero element in one of the bottom $n - r$ slots corresponding to the zero rows of A'. Hence $[A'|\mathbf{b}']$ contains a row in which the only nonzero entry lies in the last column. Thus, by the last theorem, since rank is preserved under multiplication by elementary matrices (Corollary 0.4), we have $A\mathbf{x} = \mathbf{b}$ is inconsistent iff $\operatorname{rank} A \neq \operatorname{rank}[A|\mathbf{b}]$ iff $\operatorname{rank} A' \neq \operatorname{rank}[A'|\mathbf{b}']$ iff $[A'|\mathbf{b}']$ contains a row in which the only nonzero entry lies in the last column. Conversely, if $[A'|\mathbf{b}']$ contains a row in which the only nonzero entry lies in the last column, then $\operatorname{rank}[A'|\mathbf{b}']) > \operatorname{rank}[A'|\mathbf{0}] = \operatorname{rank} A'$.

The second point follows from the previous theorem, Corollary 0.4, and 1 of this theorem: $A\mathbf{x} = \mathbf{b}$ is consistent iff $\operatorname{rank} A = \operatorname{rank} A' = \operatorname{rank}[A'|\mathbf{b}'] = \operatorname{rank}[A|\mathbf{b}]$ iff $[A'|\mathbf{b}']$ contains no row in which the only nonzero entry lies in the last column. ■

Theorem 0.11 *Let $A\mathbf{x} = \mathbf{b}$ be a system of m linear equations in n unknowns. If $B \in \mathbb{R}^{m\times m}$ is invertible, then the system $(BA)\mathbf{x} = B\mathbf{b}$ is equivalent to $A\mathbf{x} = \mathbf{b}$.*

Proof: If K is the solution set for $A\mathbf{x} = \mathbf{b}$ and K' is the solution set for $(BA)\mathbf{x} = B\mathbf{b}$, then

$$\begin{aligned} \mathbf{w} \in K &\iff A\mathbf{w} = \mathbf{b} = (B^{-1}B)\mathbf{b} \\ &\iff (BA)\mathbf{w} = B\mathbf{b} \\ &\iff \mathbf{w} \in K' \end{aligned}$$

so $K = K'$. ■

Corollary 0.12 *If $A\mathbf{x} = \mathbf{b}$ is a system of m linear equation in n unknowns, then $A'\mathbf{x} = \mathbf{b}'$ is equivalent to $A\mathbf{x} = \mathbf{b}$ if $[A'|\mathbf{b}']$ is obtained from $[A|\mathbf{b}]$ by a finite number of elementary row operations.*

Proof: If $[A'|\mathbf{b}']$ is obtained from $[A|\mathbf{b}]$ by a finite number of elementary row operations, which may be executed by left-multiplying $[A|\mathbf{b}]$ by elementary $m \times m$ matrices $E_1, \ldots, E_p$, then let $B = E_p E_{p-1} \cdots E_1$, which is invertible, so that $[A'|\mathbf{b}'] = B[A|\mathbf{b}] = [BA|B\mathbf{b}]$. Hence, since $A' = BA$ and $\mathbf{b}' = B\mathbf{b}$, $A'\mathbf{x} = \mathbf{b}'$ is equivalent to $A\mathbf{x} = \mathbf{b}$ by the previous theorem. ■

Remark 0.13 (Gaussian Elimination) *As a result of this corollary, we now know that Gaussian elimination transforms any system of linear equations $A\mathbf{x} = \mathbf{b}$ into its* **equivalent** *reduced row echelon form $A'\mathbf{x} = \mathbf{b}'$. In the* **forward pass** *the augmented matrix is transformed into an upper triangular matrix in which the first nonzero entry of each row is 1, and it occurs in a column to the right of the first nonzero entry of each preceding row. This is achieved by a finite number type 3 and 2 row operations/elementary matrix multiplications, since there are finitely many rows in $[A|\mathbf{b}]$. In the* **backward pass** *or* **back substitution** *the upper triangular matrix is transformed into reduced row echelon form by making the first nonzero entry of each row the only nonzero entry of its column. This is also achieved by type 3 and 2 row operations/elementary matrix multiplications. Hence, by the previous corollary, we can always find $m \times m$ invertible matrices B such that by multiplying the augmented matrix by it we produce an equivalent system which is in row echelon form.*

By Theorem 0.10 *through* Corollary 0.12 *we know that Gaussian elimination will tell us whether a system $A\mathbf{x} = \mathbf{b}$ does or does not have a solution, namely if and only if the reduced row echelon form of the augmented matrix $[A'|\mathbf{b}']$ contains no row in which the only nonzero entry lies in the last column. The next theorem tells us what to do next in order to obtain a particular solution $\mathbf{s}$ and, when A is not invertible, a basis for the solution set $K = \mathbf{s} + \ker(T_A)$.* ■

Theorem 0.14 *Let $A\mathbf{x} = \mathbf{b}$ be a consistent system of m linear equations in n unknowns, that is let $\operatorname{rank} A = \operatorname{rank}[A|\mathbf{b}]$, and let the reduced row echelon form $[A'|\mathbf{b}']$ of the augmented matrix $[A|\mathbf{b}]$ have $r \leq m$ nonzero rows. Then,*

(1) $\operatorname{rank} A' = r$

(2) *If we divide into two classes the variables appearing in the reduced row echelon form $A'\mathbf{x} = \mathbf{b}'$ of the system, the* **outer variables** *or* **dependent variables***, consisting of the r variables $x_1 = x_{i_1}, \ldots, x_{i_r}$ appearing as the leftmost in one of the equations, and the* **inner variables** *or* **free variables** *consisting of the other x_j, and then parametrize the inner variables $x_{j_1}, \ldots, x_{j_{n-r}}$ by setting $x_{j_1} = t_1, \ldots, x_{j_{n-r}} = t_{n-r}$ for $t_1, \ldots, t_{n-r} \in \mathbb{R}$, then, solving for the outer variables in terms of the inner variables and putting the resulting values of the x_i in terms of $t_1, \ldots, t_{n-r}$ back into the equation for $\mathbf{x}$ results in a general solution of the form*

$$\mathbf{x} = \mathbf{s} = \mathbf{s}_0 + t_1\mathbf{u}_1 + \cdots + t_{n-r}\mathbf{u}_{n-r}$$

Here, the constant vector $\mathbf{s}_0$ is a particular solution of the system, i.e. $\mathbf{s}_0 \in K$, and the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-r}\}$ is a basis for $\ker(T_A)$, the solution set to the corresponding homogeneous system. The procedure is illustrated below (cf. also Example 0.17*):*

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$$
\xrightarrow[\text{elimination}]{\text{Gaussian}}
\begin{pmatrix}
1 & a'_{12} & \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots} & a'_{1n} \\
\vdots & & & & & & \vdots \\
0 & \cdots & 0 & 1 & a'_{r,n-r+1} & \cdots & a'_{rn} \\
0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
\vdots & & & & & & \vdots \\
0 & \cdots & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} b'_1 \\ \vdots \\ b'_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

$$
\xrightarrow[\text{inner variables}]{\substack{\text{outer variables} \\ \text{in terms of}}}
\begin{cases}
x_1 = b'_1 - a'_{12}x_2 - \cdots - a'_{1n}x_n \\
x_{i_2} = b'_2 - a'_{2i_2}x_{i_2} - \cdots - a'_{1n}x_n \\
\quad \vdots \\
x_{i_r} = b'_r - a'_{r,n-r+1}x_{n-r+1} - \cdots - a'_{rn}x_{rn}
\end{cases}
$$

$$
\xrightarrow[\text{and rearranging}]{\substack{\text{parametrizing} \\ \text{the inner variables}}}
\begin{pmatrix} x_1 \\ \vdots \\ x_{j_1} \\ \vdots \\ x_{i_r} \\ \vdots \\ x_{j_{n-r}} \end{pmatrix}
=
\begin{pmatrix} b'_1 + u_{11}t_1 + \cdots + u_{1,n-r}t_{n-r} \\ \vdots \\ t_1 \\ \vdots \\ b'_r + u_{r1}t_1 + \cdots + u_{r,n-r}t_{n-r} \\ \vdots \\ t_{n-r} \end{pmatrix}
$$

the last of which may be written as a linear combination of $1, t_1, \ldots, t_{n-r}$ *and condensed to*

$$\mathbf{x} = \mathbf{s} = \mathbf{s}_0 + t_1\mathbf{u}_1 + \cdots + t_{n-r}\mathbf{u}_{n-r}$$

Proof: (1) Since $[A'|\mathbf{b}']$ is in reduced row echelon form, it must have r nonzero rows by the definition of reduced row echelon form, and they are clearly linearly independent, whence $r = \operatorname{rank}[A'|\mathbf{b}'] = \operatorname{rank} A' = r$. (2) By our methods of getting $\mathbf{s}$ we know that any such $\mathbf{s}$, for any values of $t_1, \ldots, t_{n-r}$, is a solution of the system $A'\mathbf{x} = \mathbf{b}'$, and therefore to the equivalent original system $A\mathbf{x} = \mathbf{b}$:

$$K = \mathbf{s}_0 + \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{n-r})$$

In particular, if we set $t_1 = \cdots = t_{n-r} = 0$, we see that $\mathbf{s}_0$ is a particular solution, i.e. $\mathbf{s}_0 \in K$. But by Theorem 0.7 we know that $K = \mathbf{s}_0 + \ker T_A$, whence

$$
\begin{aligned}
\ker T_A &= -\mathbf{s}_0 + K \\
&= \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{n-r})
\end{aligned}
$$

However, because $\operatorname{rank} A = r$, we must have

$$\dim(\ker T_A) = n - \operatorname{rank} T_A = n - \operatorname{rank} A = n - r$$

Therefore, we have that $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-r}\}$ is a basis for $\ker T_A$. ■

Theorem 0.15 *If* $A \in \mathbb{R}^{n\times n}$ *has rank* $r > 0$ *and* $B = \operatorname{rref}(A)$*, then*

(1) *The number of nonzero rows in* B *is* r*.*

(2) *For each* $i = 1, \ldots, r$*, there is a column* $\mathbf{b}_{j_i}$ *of* B *such that* $\mathbf{b}_{j_i} = \mathbf{e}_i$*.*

(3) *The columns of* A *numbered* $j_1, \ldots, j_r$*, corresponding to the* $\mathbf{b}_{j_i}$ *in* (2)*, are linearly independent.*

(4) *For all* $k = 1, \ldots, n$*, if column* k *of* B *is the linear combination*

$$d_1\mathbf{e}_1 + \cdots + d_r\mathbf{e}_r$$

then column k of A is the linear combination

$$d_1\mathbf{a}_{j_i} + \cdots + d_r\mathbf{a}_{j_i}$$

where the $\mathbf{a}_{j_i}$ are the linearly independent columns given in (3)*.*

Proof: (1) By Corollary 0.4 $\text{rank}(A) = r \implies \text{rank}(B) = r$, and because B is in reduced row echelon form, no nonzero row of B can be a linear combination of the others, which means B has exactly r nonzero rows. (2) If $r \geq 1$, the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ must occur among the columns of B by the definition of reduced row echelon form, and these we label $\mathbf{b}_{j_i}$. (3) Note that if there are $c_1, \ldots, c_r \in \mathbb{R}$ such that

$$c_1\mathbf{a}_{j_1} + \cdots c_r\mathbf{a}_{j_r} = \mathbf{0}$$

since B is obtained from A by a finite sequence of elementary row operations, there exists an invertible $m \times m$ matrix $C = E_pE_{p-1}\cdots E_1$ such that $CA = B$, whence

$$\begin{aligned} C(c_1\mathbf{a}_{j_1} + \cdots c_r\mathbf{a}_{j_r}) &= c_1C\mathbf{a}_{j_1} + \cdots c_rC\mathbf{a}_{j_r} \\ &= c_1\mathbf{b}_{j_1} + \cdots c_r\mathbf{b}_{j_r} \\ &= c_1\mathbf{e}_1 + \cdots c_r\mathbf{e}_r \\ &= \mathbf{0} \end{aligned}$$

so we must have $c_1 = \cdots = c_r = 0$, and $\mathbf{a}_{j_i}, \ldots, \mathbf{a}_{j_i}$ are linearly independent. (4) Note that since B has only r nonzero rows, every column of B is of the form

$$\begin{pmatrix} d_{1i} \\ \vdots \\ d_{ri} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

for $d_{1i}, \ldots, d_{ri} \in F$, so for those columns of B that look like $d_{1i}\mathbf{e}_1 + \cdots + d_{ri}\mathbf{e}_r$, the corresponding columns of A must be

$$\begin{aligned} C^{-1}(d_{1i}\mathbf{e}_1 + \cdots + d_{ri}\mathbf{e}_r) &= d_{1i}C^{-1}\mathbf{e}_1 + \cdots + d_{ri}C^{-1}\mathbf{e}_r \\ &= d_{1i}C^{-1}\mathbf{b}_{1i} + \cdots + d_{ri}C^{-1}\mathbf{b}_{ri} \\ &= d_{1i}\mathbf{a}_{1i} + \cdots + d_{ri}\mathbf{a}_{ri} \end{aligned}$$

■

Corollary 0.16 *The reduced row echelon form of a matrix is unique.*

Proof: This follows directly from part (4) of the previous theorem, since the d_{ij} are determined by the columns of A, and every column $\mathbf{a}_{j_i}$ of A is in the span of $\{\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}\}$, which is a linearly independent set. Consequently, each column is *uniquely* expressed as a linear combination of $\{\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_r}\}$, whence the d_{ij} are unique. Now, since C is invertible, T_C is an isomorphism, so that $B = T_C(A)$ also has its columns uniquely expressed as a linear combination of $T_C(\mathbf{a}_{j_1}), \ldots, T_C(\mathbf{a}_{j_r})$. Consequently, B is unique. ■

0.4 Examples

Example 0.17 *Convert the following system of 4 linear equations in 5 unknowns*

$$\begin{aligned} 2x_1 + 3x_2 + x_3 + 4x_4 - 9x_5 &= 17 \\ x_1 + x_2 + x_3 + x_4 - 3x_5 &= 6 \\ x_1 + x_2 + x_3 + 2x_4 - 5x_5 &= 8 \\ 2x_1 + 2x_2 + 2x_3 + 3x_4 - 8x_5 &= 14 \end{aligned}$$

by Gaussian elimination into reduced row echelon form, then parametrize the inner variables and show that the solution set of the system is

$$\begin{aligned} K &= (3,1,0,2,0) + \ker A \\ &= (3,1,0,2,0) + \operatorname{span}\big((-2,1,1,0,0),(2,-1,0,2,1)\big) \end{aligned}$$

Solution: We use E_1, E_2 and E_3 to denote generic elementary matrices of type 1, 2 and 3, respectively.

$$\left(\begin{array}{ccccc|c} 2 & 3 & 1 & 4 & -9 & 17 \\ 1 & 1 & 1 & 1 & -3 & 6 \\ 1 & 1 & 1 & 2 & -3 & 6 \\ 2 & 2 & 2 & 3 & -8 & 14 \end{array}\right) \xrightarrow{E_1} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -3 & 6 \\ 2 & 3 & 1 & 4 & -9 & 17 \\ 1 & 1 & 1 & 2 & -3 & 6 \\ 2 & 2 & 2 & 3 & -8 & 14 \end{array}\right) \tag{0.2}$$

$$\xrightarrow{\text{row 1: } 3\times E_3} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -3 & 6 \\ 0 & 1 & -1 & 2 & -3 & 5 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 1 & -2 & 2 \end{array}\right) \xrightarrow{\text{row 3: } E_1} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & -3 & 6 \\ 0 & 1 & -1 & 2 & -3 & 5 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{0.3}$$

$$\xrightarrow{\text{row 3: } 2\times E_3} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 0 & -1 & 4 \\ 0 & 1 & -1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\text{row 1: } E_3} \left(\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & -2 & 3 \\ 0 & 1 & -1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{0.4}$$

where (0.2) and (0.3) represent the *forward pass*, reducing $(A|\mathbf{b})$ to an upper triangular matrix with the first nonzero entry in each row equal to 1, while the *backward pass/back substitution* occurs in (0.4), producing the reduced row echelon form. The equivalent system of linear equations corresponding to the reduced row echelon matrix is

$$\begin{aligned} x_1 + \quad 2x_3 \quad - 2x_5 &= 3 \\ x_2 - x_3 \quad + x_5 &= 1 \\ x_4 - 2x_5 &= 2 \end{aligned}$$

Now, to solve such a system, divide the variables into 2 sets, one consisting of those that appear as leftmost in one of the equations of the system, the other of the rest. In this case, we divide them into $\{x_1, x_2, x_4\}$ and $\{x_3, x_5\}$. To each variable in the second set, assign a parametric value $t_1, t_2, \cdots \in \mathbb{R}$. In our case we have $x_3 = t_1$ and $x_5 = t_2$. Then solve for the variables in the first set in terms of those in the second set:

$$\begin{aligned} x_1 &= -2x_3 + 2x_5 + 3 = -2t_1 + 2t_2 + 3 \\ x_2 &= x_3 - x_5 + 1 = t_1 - t_2 + 1 \\ x_4 &= 2x_5 + 2 = 2t_2 + 2 \end{aligned}$$

Thus an arbitrary solution is of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -2t_1 + 2t_2 + 3 \\ t_1 - t_2 + 1 \\ t_1 \\ 2t_2 + 2 \\ t_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + t_1 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

Note that

$$\beta = \left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\}, \qquad \mathbf{s} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}$$

are, respectively, a basis for $\ker(T_A)$, the homogeneous system, and a particular solution of the nonhomogeneous system. Of course $\ker(T_A) = \operatorname{span}(\beta)$, so the solution set for the nonhomogeneous system is

$$\begin{aligned} K &= \mathbf{s} + \ker(T_A) \\ &= \{(3,1,0,2,0) + t_1(-2,1,1,0,0) + t_2(2,-1,0,2,1) \mid t_1, t_2 \in \mathbb{R}\} \end{aligned}$$

For example, choosing $t_1 = 2$ and $t_2 = 10$, we have $\mathbf{s} = (19, -7, 2, 22, 10)$ we have

$$\begin{pmatrix} 2 & 3 & 1 & 4 & -9 \\ 1 & 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & 2 & -5 \\ 2 & 2 & 2 & 3 & -8 \end{pmatrix} \begin{pmatrix} 19 \\ -7 \\ 2 \\ 22 \\ 10 \end{pmatrix} = \begin{pmatrix} 17 \\ 6 \\ 8 \\ 14 \end{pmatrix}$$

So $\mathbf{s}$ is indeed a solution. ■

Example 0.18 *Show that the first, third and fifth columns of*

$$A = \begin{pmatrix} 2 & 4 & 6 & 2 & 4 \\ 1 & 2 & 3 & 1 & 1 \\ 2 & 4 & 8 & 0 & 0 \\ 3 & 6 & 7 & 5 & 9 \end{pmatrix}$$

are linearly independent.

Solution: We could, of course, check directly, if we already knew that columns 1, 3 and 5 were the ones we were looking for: $\forall a, b, c \in \mathbb{R}$,

$$a \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix} + b \begin{pmatrix} 6 \\ 3 \\ 8 \\ 7 \end{pmatrix} + c \begin{pmatrix} 4 \\ 1 \\ 0 \\ 9 \end{pmatrix} = \begin{pmatrix} 2a + 6b + 4c \\ a + 3b + c \\ 2a + 8b \\ 3a + 7b + 9c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that $a = -4b$, which implies that $-8b + 6b = -4c$, or $b = 2c$, $-4b + 3b = -c$, or $b = c$, whence $c = 2c$, so $c = 0$, whence $a = b = 0$. But we might have to try $\binom{5}{3} = \frac{5!}{3!2!} = 10$ different possible combinations of columns of A to figure out that the 1, 3, 5 combination is the right one. Instead of proceeding so haphazardly, we could deduce this more simply by transforming A to reduced row echelon form and using Theorem 0.15:

$$\begin{pmatrix} 2 & 4 & 6 & 2 & 4 \\ 1 & 2 & 3 & 1 & 1 \\ 2 & 4 & 8 & 0 & 0 \\ 3 & 6 & 7 & 5 & 9 \end{pmatrix} \xrightarrow[\text{elimination}]{\text{Gaussian}} \begin{pmatrix} 1 & 2 & 0 & 4 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

which immediately shows that $\mathbf{b}_{1_1} = \mathbf{e}_1$, $\mathbf{b}_{3_2} = \mathbf{e}_2$ and $\mathbf{b}_{5_3} = \mathbf{e}_3$ are our B columns, and hence $\mathbf{a}_1$, $\mathbf{a}_3$ and $\mathbf{a}_5$ are our linearly independent A columns. Note also that since $\mathbf{b}_2 = 2\mathbf{b}_1 = 2\mathbf{e}_1$ and $\mathbf{b}_4 = 4\mathbf{b}_1 - \mathbf{b}_3 = 4\mathbf{e}_1 - \mathbf{e}_2$, we must have, by part 4 of the theorem, that $\mathbf{a}_2 = 2\mathbf{a}_1$ and $\mathbf{a}_4 = 4\mathbf{a}_1 - \mathbf{a}_3$, which we of course have. ■